Random Assignment

Thomas Wiemann
University of Chicago

Econometrics
Econ 21020

Updated: April 22, 2022

Introduction

Lecture 4 provided an introduction to causal inference:

- ▷ Used the all caused model to define potential outcomes;
- ▷ Introduced and discussed common causal parameters;
- ▷ Concluded that assumptions are necessary to learn about causal parameters from data due to the fundamental problem of causal inference;

Today: Causal inference under the Random Assignment assumption.

Assumption leveraged for identification in experimental settings:

- ▷ Laboratory experiments: E.g., in behavioral economics.
- ▷ Field experiments: E.g., in development economics.

We focus on binary (or discrete) policy variables today.

- ▷ General policy variables in Lecture 6.

Outline

1. Random Assignment
   - ▷ Definition
   - ▷ Identification of Common Causal Parameters
2. Case Study: Gneezy et al. (2019)
   - ▷ Definition of the ATE
   - ▷ ATE Identification under Random Assignment
   - ▷ Construction and Analysis of $\widehat{\text{ATE}}$
   - ▷ Hypothesis Testing
3. Evaluating Random Assignment
4. Binning Estimators

These notes benefit greatly from the lecture notes of Prof. Alex Torgovitsky.

Outline

1. **Random Assignment**
   - ▷ **Definition**
   - ▷ Identification of Common Causal Parameters
2. Case Study: Gneezy et al. (2019)
   - ▷ Definition of the ATE
   - ▷ ATE Identification under Random Assignment
   - ▷ Construction and Analysis of $\widehat{\text{ATE}}$
   - ▷ Hypothesis Testing
3. Evaluating Random Assignment
4. Binning Estimators

These notes benefit greatly from the lecture notes of Prof. Alex Torgovitsky.

The All Causes Model

We begin our analysis with the all causes model:

$$Y = g(W, U), \tag{1}$$

where

- ▷ $Y \equiv$ an outcome;
- ▷ $W \equiv$ an policy;
- ▷ $U \equiv$ all determinants of Y other than W;
- ▷ (Y, W, U) is a random vector;
- ▷ (Y, W) are observables and U are unobservables;
- ▷ and an economic model $g : \text{supp}\, W \times \text{supp}\, U \to \text{supp}\, Y$.

The potential outcomes (with the policy *fixed*) are defined as

$$g(w, U), \quad \forall w \in \text{supp}\, W. \tag{2}$$

Random Assignment

In the context of the all causes model, we may state the random assignment assumption as follows:

Assumption 1 (Random assignment; RA)

Let (Y, W, U) be a random vector with joint distribution characterized by Equation (1). *Random assignment* assumes

$$W \perp\!\!\!\perp U. \tag{3}$$

In words: the policy W is independent of all other determinants U.

- Assumption violated if (parts of) U affect the policy W.
- Most plausible in experimental settings where W is randomly and independently assigned.
- Most problematic in settings where agents make policy choices.

Random Assignment (Contd.)

Example 1

Consider the analysis of Gneezy et al. (2019), who investigate the role of student effort for their performance in the PISA test. The authors conduct a randomized control trial (RCT) where a randomly chosen subset of students receive a financial incentive to do well on the test. The idea is that financial incentives increase student effort. Here,

- ▷ $Y \equiv$ a student's PISA score;
- ▷ $W \equiv$ an indicator for having received financial incentives to do well on the PISA test;
- ▷ $U \equiv$ determinants of Y other than W (e.g., education quality).

Does RA seem plausible here?

- ▷ RA fails if students who were given the incentive were systematically different (e.g., smarter) than those who were not.
- ▷ Unless the experiment was compromised, RA is plausible by design.

***Note**: This example is due to a discussion in Prof. Alex Torgovitsky's course.*

Random Assignment (Contd.)

Example 2

Recall the returns to education example discussed in previously. Here,

- $Y \equiv$ hourly wages;
- $W \equiv$ and indicator for having obtained a college degree;
- $U \equiv$ determinants of Y other than W, e.g., intellect or connections.

Does RA seem plausible here?

- RA fails if students who obtained a college degree were systematically different than those were not.
- RA is implausible as we believe (or: hope?) that obtaining a college degree requires requires a certain level of effort / intellect.
- Students are not obtaining a college degree as if it was random: We should expect a substantial association between obtaining a college degree and socio-economic backgrounds.

Outline

1. **Random Assignment**
   - Definition
   - **Identification of Common Causal Parameters**
2. Case Study: Gneezy et al. (2019)
   - Definition of the ATE
   - ATE Identification under Random Assignment
   - Construction and Analysis of $\widehat{\text{ATE}}$
   - Hypothesis Testing
3. Evaluating Random Assignment
4. Binning Estimators

Identification

We now turn to identification of the ATE, ATT, and ATU.

For this purpose, we rely on the following result:

Corollary 1

Let (W, U) be random variables such that $W \perp\!\!\!\perp U$. Then

$$E_U[h(U)|W] = E_U[h(U)], \tag{4}$$

for all functions h.

Proof. $W \perp\!\!\!\perp U$

$\Rightarrow W \perp\!\!\!\perp h(U)$ by Corollary 2 of Lecture 2A.

$\Rightarrow E[h(U)|W] = E[h(U)]$ by Corollary 8 of Lecture 2B.

□

Identification (Contd.)

Theorem 1

Let (Y, W, U) be a random vector with joint distribution characterized by Equation (1). Under RA, the ATT and ATU are point-identified.

Proof.

1. $$ATT = E[g(1,u) - g(0,u) \mid W=1] = E[g(1,u) \mid W=1] - E[\underbrace{g(0,u)}_{\equiv h(u)} \mid W=1]$$

$$= E[g(W,u) \mid W=1] - E[g(0,u)]$$

$$= E[Y \mid W=1] - E[g(0,u) \mid W=0] = E[Y \mid W=1] - E(Y \mid W=0]$$

2. $$ATU = E[g(1,u) - g(0,u) \mid W=0] = E[g(1,u) \mid W=1] - E[g(0,u) \mid W=0]$$

$$= E[Y \mid W=1] - E[Y \mid W=0]$$

Note $ATT = ATU$.

Identification (Contd.)

Identification of the ATE follows from Theorem 1.

Corollary 2

Let (Y, W, U) be a random vector with joint distribution characterized by Equation (1). Under RA, the ATE is point-identified.

Proof.

$$ATE = E[g(1,U) - g(0,U)]$$

$$\overset{L.I.E.}{=} E[g(1,U) - g(0,U) \mid W=1] P(W=1) + E[g(1,U) - g(0,U) \mid W=0] P(W=0)$$

$$= ATT \, P(W=1) + ATU \, P(W=0) = E[Y \mid W=1] - E[Y \mid W=0]$$

□

Identification (Contd.)

Notice that under RA, we have

$$\text{ATE} = \text{ATT} = \text{ATU}. \tag{5}$$

- In words: The expected effect of the treatment for a randomly chosen individual is the same as the expected effect of the treatment for a randomly chosen treated/untreated individual.
- Since individuals are assumed to be randomly assigned to treatment under RA, this makes sense!

Assumption RA explicitly restricts selection into treatment.

- If individuals make policy decisions themselves, then likely $W \not\perp U$.
- Unfortunate because economics studies often *optimizing* agents, i.e., agents making policy decisions.

Field experiments are occasionally described as the "gold standard."

- Experiments allow for RA by design $\Rightarrow$ ATE is identified.
- But what if we are interested in selection?

Outline

1. Random Assignment
   - ▷ Definition
   - ▷ Identification of Common Causal Parameters
2. **Case Study: Gneezy et al. (2019)**
   - ▷ **Definition of the ATE**
   - ▷ ATE Identification under Random Assignment
   - ▷ Construction and Analysis of $\widehat{\text{ATE}}$
   - ▷ Hypothesis Testing
3. Evaluating Random Assignment
4. Binning Estimators

Case Study: Gneezy et al. (2019)

Let us revisit the analysis of Gneezy et al. (2019).

The authors are interested in a causal question along the lines of:

- ▷ *What is the change in the PISA score for students if they had extorted more effort?*

But measuring effort is challenging. Consider instead:

- ▷ *What is the change in the PISA score for students if they had received a financial incentive to do well?*

The idea is that financial incentives increase student efforts.

- ▷ Increase in incentives $\Rightarrow$ increase in effort.

We now complete the three tasks in the analysis of causal questions:

1. Definition of hypotheticals & a causal parameter of interest;
2. Parameter identification;
3. Parameter estimation and inference from real data;

Task 1: Definition

We begin with the all causes model. Recall (1). Here,

- ▷ $Y \equiv$ a student's PISA score;
- ▷ $W \equiv$ an indicator for having received financial incentives to do well on the PISA test;
- ▷ $U \equiv$ determinants of Y other than W (e.g., education quality).

We leave g unspecified for full generality.

A causal parameter that informs an answer to the causal question of interest is the expected effect of the financial incentive on a student's PISA score for a randomly selected student. That is,

$$\text{ATE} = E[g(1, U) - g(0, U)]. \tag{6}$$

This completes task 1 (Definition)!

Outline

1. Random Assignment
   - Definition
   - Identification of Common Causal Parameters
2. **Case Study: Gneezy et al. (2019)**
   - Definition of the ATE
   - **ATE Identification under Random Assignment**
   - Construction and Analysis of $\widehat{\text{ATE}}$
   - Hypothesis Testing
3. Evaluating Random Assignment
4. Binning Estimators

Task 2: Identification

The ATE is a function of (the distribution of) the unobservables U.

- ▷ Need an identifying assumption to (sufficiently) restrict the joint distribution of (Y, W, U).

Gneezy et al. (2019) assign incentives to a random subset of students.

- ▷ RA is plausible by design.

We continue our analysis with assuming RA.

- ▷ Under RA, the ATE is point identified by Corollary 2.

This completes task 2 (Identification)!

Outline

1. Random Assignment
   - ▷ Definition
   - ▷ Identification of Common Causal Parameters
2. **Case Study: Gneezy et al. (2019)**
   - ▷ Definition of the ATE
   - ▷ ATE Identification under Random Assignment
   - ▷ **Construction and Analysis of $\widehat{\text{ATE}}$**
   - ▷ Hypothesis Testing
3. Evaluating Random Assignment
4. Binning Estimators

Task 3: Estimation

Task 3 (Estimation & Inference) will require two key steps:

- ▷ Construct a useful estimator of the ATE under RA.
- ▷ Characterize its sampling distribution.

Fortunately, our identification proof of the ATE under RA was constructive. In particular, we showed that we can write the ATE as

$$\text{ATE} = \mu_{Y|1} - \mu_{Y|0}, \tag{7}$$

where $\mu_{Y|w} \equiv E[Y|W = w]$ for $w \in \{0, 1\}$.

This is just a difference in two familiar conditional expectations:

- ▷ Problem Set 2 constructed and analyzed an estimator $\hat{\mu}_{Y|1}$ for $\mu_{Y|1}$;
- ▷ TA Session 3 constructed and analyzed an estimator $\hat{\mu}_{Y|0}$ for $\mu_{Y|0}$.

The sample analogue estimate of the ATE is then simply given by

$$\widehat{\text{ATE}} = \hat{\mu}_{Y|1} - \hat{\mu}_{Y|0} = \frac{\frac{1}{n}\sum Y_i W_i}{\frac{1}{n}\sum W_i} - \frac{\frac{1}{n}\sum Y_i(1-W_i)}{\frac{1}{n}\sum(1-W_i)} \tag{8}$$

Task 3: Estimation (Contd.)

Suppose we observe a sample $(Y_1, W_1), \dots, (Y_n, W_n) \overset{iid}{\sim} (Y, W)$.

Our previous analysis of $\hat{\mu}_{Y|1}$ and $\hat{\mu}_{Y|0}$ showed that

$$\frac{\left(\hat{\mu}_{Y|w} - E[Y|W = w]\right)}{se(\hat{\mu}_{Y|w})} \xrightarrow{d} N(0, 1), \qquad w = 0, 1, \tag{9}$$

where

$$se(\hat{\mu}_{Y|w}) = \frac{\hat{\sigma}_{Y|w}}{\sqrt{n}\sqrt{\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_w(W_i)}}, \tag{10}$$

with $\hat{\sigma}_{Y|w}$ being the sample analogue estimator for $sd(Y|W = w)$.

- ▷ Characterizes the *marginal* sampling distribution of $\hat{\mu}_{Y|1}$ and $\hat{\mu}_{Y|0}$.
- ▷ Now need the *joint* sampling distribution of $\hat{\mu}_{Y|1}$ and $\hat{\mu}_{Y|0}$.

Task 3: Estimation (Contd.)

We proceed in steps:

1. Show that

$$\sqrt{n}\left(\begin{bmatrix}\hat{\mu}_{Y|1}\\ \hat{\mu}_{Y|0}\end{bmatrix} - \begin{bmatrix}\mu_{Y|1}\\ \mu_{Y|0}\end{bmatrix}\right) \xrightarrow{d} N(0, \Sigma).$$

2. Derive expression for Σ in terms of expectations of (Y, W).

3. Find t and σ^2_{ATE} such that Use

$$\sqrt{n}t^\top\left(\begin{bmatrix}\hat{\mu}_{Y|1}\\ \hat{\mu}_{Y|0}\end{bmatrix} - \begin{bmatrix}\mu_{Y|1}\\ \mu_{Y|0}\end{bmatrix}\right) = \sqrt{n}\left(\widehat{\text{ATE}} - \text{ATE}\right) \xrightarrow{d} N\left(0, \sigma^2_{\text{ATE}}\right)$$

4. Construct a consistent estimate $\hat{\sigma}^2_{\text{ATE}}$ for σ^2_{ATE}.

5. Show that

$$\hat{\sigma}^{-1}_{\text{ATE}}\sqrt{n}\left(\widehat{\text{ATE}} - \text{ATE}\right) \xrightarrow{d} N(0, 1).$$

Task 3: Estimation (Contd.)

1. Show that

$$\sqrt{n}\left(\begin{bmatrix}\hat{\mu}_{Y|1}\\ \hat{\mu}_{Y|0}\end{bmatrix} - \begin{bmatrix}\mu_{Y|1}\\ \mu_{Y|0}\end{bmatrix}\right) \xrightarrow{d} N(0,\Sigma).$$

$U_i \equiv Y_i - E[Y_i|W_i]$

$$\sqrt{n}\left(\begin{bmatrix}\frac{\sum Y_i W_i}{\sum W_i}\\ \frac{\sum Y_i(1-W_i)}{\sum(1-W_i)}\end{bmatrix} - \begin{bmatrix}\mu_{Y|1}\\ \mu_{Y|0}\end{bmatrix}\right) = \sqrt{n}\begin{bmatrix}\frac{\sum U_i W_i}{\sum W_i}\\ \frac{\sum U_i(1-W_i)}{\sum(1-W_i)}\end{bmatrix} = \begin{bmatrix}(\frac{1}{n}\sum W_i)^{-1} & 0\\ 0 & (\frac{1}{n}\sum(1-W_i))^{-1}\end{bmatrix}\sqrt{n}\begin{bmatrix}\frac{1}{n}\sum U_i W_i\\ \frac{1}{n}\sum U_i(1-W_i)\end{bmatrix}$$

CLT + WLLN + CMT + Slutsky,

$$\xrightarrow{d} \begin{bmatrix}E[W]^{-1} & 0\\ 0 & E[1-W]^{-1}\end{bmatrix} N(0,\Sigma), \quad \Sigma = \begin{bmatrix}Var(UW) & Cov(UW, U(1-W))\\ Cov(UW, U(1-W)) & Var(U(1-W))\end{bmatrix}$$

Theorem 7 in 3A.

Task 3: Estimation (Contd.)

2. Derive expression for Σ in terms of expectations of (Y, W).

By PSet 2: $Var(uW) = Var(Y|W=1)P(W=1)$

By TA Session 3: $Var(u(1-W)) = Var(Y|W=0)P(W=0)$

Only thing we need: off-diagonals!

$$Cov(uW, u(1-W)) = E[\underbrace{u^2 W(1-W)}_{=0}] - \underbrace{E[uW]}_{=0}\underbrace{E[u(1-W)]}_{=0} = 0$$

Hence, $\sqrt{n}\left(\begin{bmatrix}\hat{\mu}_{Y|1}\\ \hat{\mu}_{Y|0}\end{bmatrix} - \begin{bmatrix}\mu_{Y|1}\\ \mu_{Y|0}\end{bmatrix}\right) \xrightarrow{d} \begin{bmatrix} E[W]^{-1} & 0 \\ 0 & E[1-W]^{-1}\end{bmatrix} \mathcal{N}(0, \Sigma)$

$$\stackrel{d}{=} \mathcal{N}\left(0, \begin{bmatrix} \frac{Var(Y|W=1)}{P(W=1)} & 0 \\ 0 & \frac{Var(Y|W=0)}{P(W=0)} \end{bmatrix}\right)$$

Task 3: Estimation (Contd.)

3. Find t and σ^2_{ATE} such that Use

$$\sqrt{n}t^\top \left(\begin{bmatrix} \hat{\mu}_{Y|1} \\ \hat{\mu}_{Y|0} \end{bmatrix} - \begin{bmatrix} \mu_{Y|1} \\ \mu_{Y|0} \end{bmatrix} \right) = \sqrt{n} \left(\widehat{\text{ATE}} - \text{ATE} \right) \xrightarrow{d} N\left(0, \sigma^2_{\text{ATE}}\right)$$

$$ATE = \mu_{Y|1} - \mu_{Y|0} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}^\top \begin{bmatrix} \mu_{Y|1} \\ \mu_{Y|0} \end{bmatrix}, \quad t = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

By Slutsky's

$$\sqrt{n}\, t^\top \left(\begin{bmatrix} \hat{\mu}_{Y|1} \\ \hat{\mu}_{Y|0} \end{bmatrix} - \begin{bmatrix} \mu_{Y|1} \\ \mu_{Y|0} \end{bmatrix} \right) \xrightarrow{d} N\left(0, t^\top \begin{bmatrix} \frac{Var(Y|W=1)}{P(W=1)} & 0 \\ 0 & \frac{Var(Y|W=0)}{P(W=0)} \end{bmatrix} t \right)$$

$$\stackrel{d}{=} N\left(0, \underbrace{\frac{Var(Y|W=1)}{P(W=1)} + \frac{Var(Y|W=0)}{P(W=0)}}_{\equiv \sigma^2_{ATE}} \right)$$

Task 3: Estimation (Contd.)

4. Construct a consistent estimate $\hat{\sigma}^2_{\text{ATE}}$ for σ^2_{ATE}.

$$\hat{\sigma}^2_{ATE} = \frac{\hat{\sigma}^2_{Y|1}}{\hat{p}_{W=1}} + \frac{\hat{\sigma}^2_{Y|0}}{\hat{p}_{W=0}} \xrightarrow{p} \sigma^2_{ATE} \quad \text{by CMT.}$$

$$\frac{\hat{\sigma}^2_{Y|1}}{\hat{p}_{W=1}} \xrightarrow{p} \frac{\sigma^2_{Y|1}}{p(W=1)} \quad \text{by PSet 2}$$

$$\frac{\hat{\sigma}^2_{Y|0}}{\hat{p}_{W=0}} \xrightarrow{p} \frac{\sigma^2_{Y|0}}{p(W=0)} \quad \text{by TA Session 2}$$

Task 3: Estimation (Contd.)

5. Show that

$$\hat{\sigma}_{\text{ATE}}^{-1}\sqrt{n}\left(\widehat{\text{ATE}} - \text{ATE}\right) \xrightarrow{d} N(0,1).$$

$\sqrt{n}\left(\widehat{\text{ATE}} - \text{ATE}\right) \xrightarrow{d} N(0, \sigma^2_{ATE})$ by 3.

Note: 1. $A_n \equiv \hat{\sigma}^2_{ATE}$

2. $g(a) = 1/\sqrt{a}$

3. By 4., $\hat{\sigma}^2_{ATE} \xrightarrow{p} \sigma^2_{ATE}$

4. By CMT, $g(\hat{\sigma}^2_{ATE}) \xrightarrow{p} \sigma^{-1}_{ATE}$, $\forall \sigma^2_{ATE} > 0$.

Hence, by Slutsky's $\hat{\sigma}^{-1}_{ATE}\sqrt{n}(\widehat{ATE} - ATE) \xrightarrow{d} N(0,1)$.

Task 3: Estimation (Contd.)

Our analysis implies that for large n, we may approximate the sampling distribution of $\widehat{\text{ATE}}$ by a normal distribution. I.e.,

$$\widehat{\text{ATE}} \overset{d}{\approx} N\left(\text{ATE},\ se\left(\widehat{\text{ATE}}\right)\right), \tag{11}$$

where

$$se\left(\widehat{\text{ATE}}\right) = \frac{\hat{\sigma}_{\text{ATE}}}{\sqrt{n}}. \tag{12}$$

For $\alpha \in (0,1)$, we may thus construct a $1-\alpha$ confidence interval by

$$C_n = \left[\widehat{\text{ATE}} - \Phi^{-1}\left(1-\frac{\alpha}{2}\right) se\left(\widehat{\text{ATE}}\right),\ \widehat{\text{ATE}} + \Phi^{-1}\left(1-\frac{\alpha}{2}\right) se\left(\widehat{\text{ATE}}\right)\right] \tag{13}$$

This completes the theoretical analysis: Now we need to implement it!

ATE Estimation in R

ATE Estimation in R under Assumption RA

```
# Find treated and untreated individuals
y_1 <- y[w == 1]
y_0 <- y[w == 0]

# Compute conditional averages
mu_1 <- mean(y_1)
mu_0 <- mean(y_0)

# Compute standard error
n <- length(y)
p_1 <- mean(w == 1)
p_0 <- mean(w == 0)
se <- sqrt((var(y_1) / p_1 + var(y_0) / p_0) / n)

# Compute confidence set for pre-defined alpha
c_alpha <- qnorm(1 - alpha / 2)
conf <- mu_1 - mu_0 + c(-1, 1) * c_alpha * se
```

Estimation Results

Using the data from Gneezy et al. (2019), we can compute

$$\widehat{\text{ATE}} = 0.620$$

and

$$se\left(\widehat{\text{ATE}}\right) = 0.384$$

A 95% confidence interval is thus given by

$$c_n = [-0.132, 1.372]$$

Note: *On Canvas, you can find the datafile* `pisa19.csv` *used for calculating the estimates. The corresponding* R *script is on GitHub:* `example_pisa19`.

Outline

1. Random Assignment
   - ▷ Definition
   - ▷ Identification of Common Causal Parameters
2. **Case Study: Gneezy et al. (2019)**
   - ▷ Definition of the ATE
   - ▷ ATE Identification under Random Assignment
   - ▷ Construction and Analysis of $\widehat{\text{ATE}}$
   - ▷ **Hypothesis Testing**
3. Evaluating Random Assignment
4. Binning Estimators

Hypothesis Testing

Suppose now that we are worried about the incentives having a negative effect, on average. We may then consider a test where

$$H_0 : \text{ATE} \leq 0 \quad \text{versus} \quad H_1 : \text{ATE} > 0.$$

We can construct a test statistic as discussed in Lecture 3B. In particular,

$$T_n = \frac{(\widehat{ATE} - 0)}{se(\widehat{ATE})}$$

Instead of pre-specifying a α, we decide to calculate a p-value via

$$p\text{-value} = 1 - \Phi(T_n)$$

Hypothesis Testing (Contd.)

We implement the hypothesis text in R using the following code:

Testing H_0 : ATE $\leq$ `ate_0` versus H_1 : ATE $>$ `ate_0` in R

```
# Compute test-statistic
Tn <- (mu_1 - mu_0 - ate_0) / se
# Compute p-value
1 - pnorm(Tn)
```

Using the data of Gneezy et al. (2019), we compute

$$T_n = 1.616 \quad , \text{ and } \quad p\text{-value} = 0.053$$

Therefore, on a 5% significance level, we fail to reject H_0 : ATE ≤ 0. What's the correct interpretation?

- ▷ "~~We are 95% sure that incentivizing students worsened their score~~."
- ▷ "There is insufficient evidence to reject that incentivizing students worsened their score on a 5% significance level."

Outline

1. Random Assignment
   - ▷ Definition
   - ▷ Identification of Common Causal Parameters
2. Case Study: Gneezy et al. (2019)
   - ▷ Definition of the ATE
   - ▷ ATE Identification under Random Assignment
   - ▷ Construction and Analysis of $\widehat{\text{ATE}}$
   - ▷ Hypothesis Testing
3. **Evaluating Random Assignment**
4. Binning Estimators

Evaluating Random Assignment

RA is a simple but very strong restriction on the joint of (Y, W, U).

- ▷ Experiments (or RCTs) are designed to ensure RA is satisfied.
- ▷ But: "Vertrauen ist gut, Kontrolle ist besser." (German proverb)
- ▷ Researchers need to convince their audience that RA is plausible.

Given the strength of the RA assumption, it's worth contemplating whether we can check it's plausibility using data.

- ▷ Because the sampling process provides no information on the entirety of U, it's impossible for us to verify RA.
- ▷ But RA has implications that we can test. If these implications implications don't hold, this may suggest we should question whether RA holds.
- ▷ In practice: "The absence of bad news is good news."

Balance Tests

Recall that RA assumes $W \perp\!\!\!\perp U$.

Now suppose that we observe *some* variables in U, say X.

- X commonly referred to as *covariates.*
- Covariates are not policy variables of interest, but are observables.
- RA implies $W \perp\!\!\!\perp X$.

Since both are observable, we may construct a test! In particular, $W \perp\!\!\!\perp X$ implies $E[X|W = 1] = E[X|W = 0]$. Then we could test whether

$$H_0 : \mu_{X|1} = \mu_{X|0} \quad \text{versus} \quad H_1 : \mu_{X|1} \neq \mu_{X|0},$$

where $\mu_{X|1} \equiv E[X|W = 1]$ and $\mu_{X|0} \equiv E[X|W = 0]$.

This test is referred to as a *balance test.*

- Assesses whether individuals with different characteristics (X) are randomly distributed across treated and untreated groups.

Balance Tests (Contd.)

Suppose we observe $(Y_1, W_1, X_1), \ldots, (Y_n, W_n, X_n) \overset{iid}{\sim} (Y, W, X)$.

Do we now have to redo our asymptotic analysis?

- Fortunately: No! It's a comparison of conditional means as before.
- Essentially just as before but with X instead of Y.

Let $\hat{\mu}_{X|1}$ and $\hat{\mu}_{X|0}$ denote sample analogue estimators for $\mu_{X|1}$ and $\mu_{X|0}$, respectively. Then we can construct a test-statistic for a two-sided test:

$$T_n = \left| \frac{\hat{\mu}_{X|1} - \hat{\mu}_{X|0}}{se(\hat{\mu}_{X|1} - \hat{\mu}_{X|0})} \right|$$

The corresponding p-value is given by

$$p\text{-value} = 2(1 - \Phi(T_n))$$

What does it mean to reject the balancing test?

- Should raise eyebrows: Variables should be $\approx$ equal across groups.
- But not perfect: Type I errors can't be ruled-out.

Balance Tests (Contd.)

Example 3

Let's revisit Gneezy et al. (2019). Let the X denote a student's age. We compute

$$\hat{\mu}_{X|1} - \hat{\mu}_{X|0} = -0.083 \quad , \text{ and } \quad se\left(\hat{\mu}_{X|1} - \hat{\mu}_{X|0}\right) = 0.033$$

And also

$$T_n = 2.512 \quad , \text{ and } \quad p\text{-value} = 0.012$$

Therefore, on a 5% significance level, we reject $H_0 : \mu_{X|1} = \mu_{X|0}$.

- "There is sufficient evidence to reject that incentivized and un-incentivized students have the same expected age on a 5% significance level."
- Incentivized students are about 1 month younger than non-incentivized students. Doesn't raise confidence hugely...
- ... but also doesn't *prove* that RA is violated: Type I errors exist!

Outline

1. Random Assignment
   - ▷ Definition
   - ▷ Identification of Common Causal Parameters
2. Case Study: Gneezy et al. (2019)
   - ▷ Definition of the ATE
   - ▷ ATE Identification under Random Assignment
   - ▷ Construction and Analysis of $\widehat{\text{ATE}}$
   - ▷ Hypothesis Testing
3. Evaluating Random Assignment
4. **Binning Estimators**

Binning Estimators

All of the estimators considers thus far are examples of *binning estimators*. For a random vector (Y, W), a binning estimator for

$$\mu_{Y|w} \equiv E[Y|W = w], \tag{14}$$

is given by

$$\hat{\mu}_{Y|w} \equiv \frac{\sum_i^n Y_i \mathbb{1}_w(W_i)}{\sum_i^n \mathbb{1}_w(W_i)}, \tag{15}$$

$\forall w \in \operatorname{supp} W$.

By defining $X_w \equiv \mathbb{1}_w(W)$, the asymptotic analysis from Problem Set 2 applies, so that you have already derived it's asymptotic distribution...

... at least for the case when $P(W = w) \neq 0$.

- ▷ Recall that for discrete W, $w \in \operatorname{supp} W \Leftrightarrow P(W = w) > 0$;
- ▷ But for continuous W, $P(W = w) = 0, \forall w \in \operatorname{supp}$.

Binning Estimators (Contd.)

We thus *need* an alternative estimator when W is not discrete.

Note that even when W is discrete, estimation variance may be very large if $P(W = w)$ is very small (but not zero), since we showed that

$$\sqrt{n}\left(\hat{\mu}_{Y|w} - \mu_{Y|w}\right) \overset{d}{\to} N\left(0, \frac{Var(Y|W = w)}{P(W = w)}\right), \tag{16}$$

with a division by a very small number to obtain the asymptotic variance.

This is known as the *small bin problem*. Binning estimator not versatile.

- Unfortunate because you showed in problem set 2 that

$$E\left[\hat{\mu}_{Y|w} \middle| \sum_{i=1}^{n} \mathbb{1}_w(W_i) > 0\right] = \mu_{Y|w}.$$

- May prefer some bias if variance is lower: A bias-variance trade-off!

We thus *want* an alternative estimator when $P(W = w)$ is small.

Summary

We've completed our first complete causal analysis today!

- ▷ Used the all causes model to define the ATE, which we deemed informative for the causal question of interest;
- ▷ Assumed Random Assignment to prove identification of the ATE;
- ▷ Constructed, analyzed, and computed an estimator of the ATE;

In the process, we stumbled upon a statistical difficulty:

- ▷ The binning estimator is infeasible for non-discrete policies...
- ▷ ... and potentially undesirable even for discrete policies.

In the next lecture, we introduce simple linear regression to construct estimates of causal parameters (under RA) when the binning estimator is infeasible or undesirable.

- ▷ Linear regressions is *the* estimator in applied economics.
- ▷ Very convenient statistical tool, but challenging to interpret.

References

Gneezy, U., List, J. A., Livingston, J. A., Qin, X., Sadoff, S., and Xu, Y. (2019). Measuring success in education: the role of effort on the test itself. *American Economic Review: Insights*, 1(3):291–308.